1986

The Inter-Rater Reliability of the American Association on Mental Deficiency Adaptive Behavior Scale (School Edition) Using Secondary Special Education Personnel as Raters

Jimmy L. Camp Jr.
Eastern Illinois University

THESIS REPRODUCTION CERTIFICATE

TO: Graduate Degree Candidates who have written formal theses.

SUBJECT: Permission to reproduce theses.

The University Library is receiving a number of requests from other institutions asking permission to reproduce dissertations for inclusion in their library holdings. Although no copyright laws are involved, we feel that professional courtesy demands that permission be obtained from the author before we allow theses to be copied.

Please sign one of the following statements:

Booth Library of Eastern Illinois University has my permission to lend my thesis to a reputable college or university for the purpose of copying it for inclusion in that institution's library or research holdings.

I respectfully request Booth Library of Eastern Illinois University not allow my thesis be reproduced because _____ ____

_____

_____

_____ _____

Date Author

m

The Inter-Rater Reliability of the American Association on Mental Deficiency Adaptive Behavior Scale (School Edition) Using Secondary Special Education Personnel as Raters

(TITLE)

BY

Jimmy L. Camp, Jr.

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

Master of Science in Education

IN THE GRADUATE SCHOOL, EASTERN ILLINOIS UNIVERSITY
CHARLESTON, ILLINOIS

1986
YEAR

I HEREBY RECOMMEND THIS THESIS BE ACCEPTED AS FULFILLING
THIS PART OF THE GRADUATE DEGREE CITED ABOVE

Abstract

The concept of adaptive behavior has been present in the definition of adaptive behavior for many years. With the advent of intelligence testing, adaptive behavior was deemphasized, but later reemerged as widespread dissatisfaction with the testing movement grew among researchers in the area of mental retardation. The importance of the concept of adaptive behavior has been affirmed by the American Association on Mental Deficiency (A.A.M.D.), which developed a dual criteria system of defining mental retardation that included adaptive behavior as a major component. The Federal government later adopted this classification system for funding purposes in P.L. 94-142.

The valid and reliable measurement of adaptive behavior is a major concern of researchers, diagnosticians, and teachers in the field of mental retardation. The present study investigated the inter-rater reliability of one of the most widely used measures of adaptive behavior: the A.A.M.D. Adaptive Behavior Scale (School Edition). A total of 18 subjects (9 male and 9 female) were randomly assigned to 2 raters. The raters used the ABS-SE to rate the adaptive behavior of the subjects. The ratings of the scorers were then compared using a simple agreement formula and

kappa (a correlational statistic which accounts for chance agreements).

The results of this study show that: (1) based on total factor scores the ABS-SE may be useful in making placement decisions; (2) the ABS-SE may provide useful information for making some programming decisions; and (3) the contention of the authors, that teachers provide the most reliable information, when the first person method of assessment is used, was supported. Comments by the investigators and raters provide suggestions that may increase the reliability of the ABS-SE. Several questions that need further investigation are also raised by the author.

This investigation provides the impetus for further investigation of the inter-rater reliability of the ABS-SE. It also provides valuable information as to the amount of confidence educators and diagnosticians should place in this popular instrument. This study has shown that those concerned with the reliable and valid measurement of adaptive behavior must continuously refine the instruments used to assess this concept.

Dedication

This work is dedicated to my Lord and Savior: Jesus Christ, who has given me inspiration and comfort during the trying times and called me, through the Holy Spirit, to pursue a vocation of the highest honor: teaching.

This work is also dedicated to my Grandfather, Rev. George W. Camp, and my Grandmother, Marie L. Camp, whose love of life and love of God have made me the person I am today.

Acknowledgements

I would like to thank the members of my thesis committee: Dr. Andrew Brulle, Dr. Kathlene Shank, and Dr. Mark Davis, for their guidance, critiques, and time during the various stages of this investigation. A special word of thanks goes to Dr. Brulle, who has been more than just an advisor, a professor, or a thesis chairperson; he has been an inspiration to achieve the highest ideals of the profession of Education. The entire Department of Special Education at Eastern Illinois University deserve credit for their academic and emotional support of this work.

A special thank you goes to the Special Education teachers and special needs counselor of Danville High School in Danville, Illinois for their participation as raters in this investigation.

Words cannot describe the debt of gratitude that I owe to all those persons who supported me emotionally during graduate school. My greatest thanks goes to Jackie Holt, Lucy Garlock, Kathy Kirby, and Drew Hoffman; who were more than my friends; they were family. I would also like to thank a very special friend who helped me "survive" the thesis writing process: Wendy Barnes.

The Inter-Rater Reliability of the American Association on Mental Deficiency Adaptive Behavior Scale (School Edition) Using Secondary Special Education Personnel as Raters

The concept of adaptive behavior has been present in the definition of mental retardation for many years. Such terms as social competency, skills training, the power of fending for oneself in life, adaptability to environment, and efficiency of social value have all been described as components in the 19th and 20th century medical definitions of mental retardation. These components are now known as adaptive behavior (Lambert, Windmiller, Cole & Figueroa, 1975).

Adaptive behavior, though present in the early definitions of mental retardation, was deemphasized with the advent of intelligence testing. The years between 1905 (the beginning of Binet's work for the Paris schools) and 1959 (the American Association on Mental Deficiency (A.A.M.D.) establishment of the dual component classification system of mental retardation) were literally the "Dark Ages" of the concept of adaptive behavior in the field of mental retardation (Grossman, 1973; Scheerenberger, 1983). Goodenough (1949) describes the common practices of the era by stating:

> Mental ages and IQ scores from half a dozen different group tests were joyfully compiled and entered on children's permanent record cards by teachers and principals with as much assurance as their grandfathers had placed in the skull maps drawn up by their favorite phrenologist. The decade of the twenties was the heydey of the testing movement, the age of innocence when an IQ was an IQ and few ventured to doubt its omnipotence (p. 68).

The testing movement was not without its critics. Men such as Kulhman, Porteus, and Doll were outspoken opponents of a purely psychomatic approach to defining mental retardation (Scheerenberger, 1983). Perhaps the most vocal opponent to the use of psychometric testing as the sole basis for classifying individuals as mentally retarded was Edgar Doll (Scheerenberger, 1983). Doll (1924) raised several issues that he felt must be resolved before even the slightest amount of confidence could be placed in the use of intelligence tests. He called for the investigation of the effects that such factors as race, nationality, age and emotion might have on the reliability and validity of psychometric testing. He urged the adoption of uniform methods of standardization, calculation, and interpretation of

intelligence tests and their results. Doll suggested that even with low scores on intelligence tests, some people still managed to compensate for their disabilities and become productive members of society. He felt that this was a phenomenon that warranted further investigation. Doll did not feel that the testing movement was without merit, but that caution should be used in granting a single instrument so much control over the lives of human beings. He emphasizes this point by stating:

> . . . uncertainty need not prevent us from our practical service, but should certainly encourage us to proceed cautiously and with due respect for the rapid progress which is being made in the academic field (Doll, 1924, p. 360).

These charges, plus those of others, such as Yepsen (1941), who challenged the American Association on Mental Deficiency to resolve the many social, educational, legal, and research problems for the good of society and the individual labeled mentally retarded, were the impetus for the "rediscovery" of the concept of adaptive behavior (Scheerenberger, 1983).

Doll and other researchers at the Vineland School were the foremost authorities on the assessment of the resurrected concept of adaptive behavior. Porteus (1921) developed a social rating scale which identified persons as mildly mentally retarded as those persons who exhibited a lack of planning capacity, suggestability, obtrusiveness, and moodiness. Doll (1929) developed a scorecard for evaluating behavior and changes in behavior of mental patients' behavior before, during, and after psychiatric care. In 1936, Doll developed the Vineland Social Maturity Scale. This 117 item rating scale would be the standard instrument used for the measurement of adaptive behavior over the next 40 years.

The literature indicates that the concept of adaptive behavior, though present in the early medical definitions of mental retardation, was deemphasized with the advent of intelligence testing. It then reappears when researchers begin to raise questions about the reliability and validity of psychometric testing. This "swinging of the pendulum" indicates that even though the testing movement was strongly supported by researchers and educators, the concept of adaptive behavior was more strongly imbedded into the philosophies and ideologies of those who worked with

individuals identified as mentally retarded. The reemergence of the concept of adaptive behavior in the field of mental retardation did not occur immediately, nor did it occur without painstaking research on the part of many people.

In 1959, a committee of the American Association on Mental Deficiency (A.A.M.D.), chaired by Dr. Rick Herber, made adaptive behavior a major component of its definition of mental retardation. The committee defined mental retardation as "subaverage general intellectual functioning which originates during the developmental period and is associated with impairments in adaptive behavior" (Herber, 1961, p. 1). Adaptive behavior was described as having two major components: "(a) the degree to which the individual is able to function and maintain himself independently, and (b) the degree to which he meets satisfactorily the culturally imposed demands of personal and social responsibility" (Herber, 1961, p. 61). Scheerenberger (1983) identifies widespread dissatisfaction with the current testing movement combined with the mental hygiene movement and an evolution in the concept of mental retardation as the three factors that led to a dual criteria definition of mental retardation.

Although a dual classification system was adopted by the A.A.M.D., educators and diagnosticians still continued to use IQ as a singular criteria for placement when state legislatures began providing funds for the education of students labeled mentally retarded (Garrison, 1960; Lambert, Windmiller, Cole, & Figueroa, 1975; Lord, 1956; Scheerenberger, 1983). This practice drew criticism from many different groups associated with research and service delivery in the field of mental retardation. Researchers (Dunn, 1968; Garrison, 1960; Lord, 1956) and the Presidents's Panel on Mental Retardation (1962) attacked the practice of placement into special education programs based solely on an IQ score. A number of legal battles also ensued following this practice of making placement decisions. Such court decisions as Larry P. v. Wilson Riles (1973) and Diana v. California State Board of Education (1970) prohibited the placement of minority students in a special education program solely on the basis of an IQ score.

In 1973, the A.A.M.D. again revised its definition of mental retardation; this time to ensure that both intelligence and adaptive behavior data were gathered and considered in the placement of students into programs for individuals labeled mentally retarded.

Under the direction of Dr. Grossman, an A.A.M.D. committee developed the following definition of mental retardation: "significantly subaverage general intellectual functioning existing concurrently with deficits in adaptive behavior, manifested in the developmental period" (p. 11). The most important difference between the definition developed by Herber (1961) and its revision by Grossman (1973) is that the relationship between adaptive behavior and intelligence for the classification of persons as mentally retarded was strengthened (Payne, Polloway, Smith, & Payne, 1981). The 1973 definition was as the official definition for funding purposes the the federal government in P.L. 94-142.

The concept of adaptive behavior after its reemergence in the field of mental retardation was slowly reaffirmed as a major tenet in the identification, placement and training of individuals labeled mentally retarded. Widespread support of a dual criteria system of defining mental retardation led to the development of such a definition by the A.A.M.D. in 1959 and its revision in 1973. perhaps the greatest affirmation of the importance of adaptive behavior was the adoption of the A.A.M.D. definition of mental

retardation (Grossman 1973) in P.L. 94-142. With the establishment of adaptive behavior as a major component of mental retardation there arose a new problem: how to measure adaptive behavior in a valid and reliable manner.

Measures of Adaptive Behavior

As previously mentioned, the Vineland Social Maturity Scale (VSMS) (Doll, 1953) was the standard instrument used for the assessment of adaptive behavior for many years. The instrument consisted of 117 items grouped into the areas of Self-Help (Dressing), Self-Help (Eating), Self-Help (General), Locomotion, Occupation, Communication, Self-Direction, and Socialization. The VSMS is administered in an interview format to a third person who is familiar with the behavior of the person being rated using the scale. The norming population consisted of 620 white subjects with 10 males and 10 females being assigned to each age level (birth to 30 years). This standardization project took place in New Jersey in 1935. The technical adequacy of the instrument has also been questioned. Salvia and Ysseldyke (1978) state "The VSMS is a venerable instrument used to assess social competence. We believe it is badly in need of revision and updating. The item

placement may no longer be appropriate and the sample is very restricted" (p. 435). The Vineland Social Maturity Scale has recently been revised and its use remains widespread throughout the nation (Salvia & Ysseldyke, 1985).

The revision of the Vineland Social Maturity Scale (renamed the Vineland Adaptive Behavior Scales by its authors) is organized around four Behavior Domains: Communication, Daily Living Skills, Socialization and Motor Skills. The use of the professional interviewer/respondent format and the requirements that a subject actually perform a described behavior on a regular basis in order to receive full item credit are characteristics which have been retained from the VSMS (Campbell, 1985). Two forms of the Vineland Adaptive Behavior Scales (VABS) are available for use by diagnosticians: the Survey Form and the Expanded Form.

Campbell (1985) examined the technical adequacy of the VABS and made the following comments:

> Adaptive behavior assessment, by means of the VABS Survey Form or the Expanded Form, can be well accomplished for the 0 to 19 year range on which the Survey Form was standardized. Older MR adults can also be evaluated although raw score conversions for these subjects are

> not expressed in standard scores, but percentiles only. The development of remediation plans, intended by the authors to be a special use of the expanded form, is subject to limitations at certain age ranges and at certain classes of behaviors. The major caution to be emphasized would seem to be that the structure, recommended uses, and methods of developing derived scores are quite different between the two forms. The potential user, then, is advised to know the properties of each quite well and fit those properties to the assessment needs of a specific client (p. 1661).

Development of the A.A.M.D. Adaptive Behavior Scales

In the late 1960's a task force under the direction of Dr. Henry Leland was appointed by the American Association on Mental Deficiency, the National Institute on Mental Health, and the Kansas Division of Institutional Management to study the assessment of adaptive behavior (Clausen, 1972). The task force reviewed all instruments purporting to measure concepts related to adaptive behavior in the United States and England and developed a number of items and scales on personal independence. They also undertook surveys of

social expectations and critical incidents in community of institutional settings to derive content areas for the maladaptive behavior scales (Nihira & Nihira, 1975; Spreat, 1982). The result of this research project was the American Association on Mental Deficiency Behavior Checklist (Nihira, Foster, Shellhaas, & Leland, 1967).

This checklist consisted of 10 domains of personal and social functioning and 12 domains of maladaptive behavior. The checklist was later revised twice and renamed the American Association on Mental Deficiency Adaptive Behavior Scale (ABS) (Nihira, Foster, Shellhaas, & Leland, 1969; Nihira, Foster Shellhaas, & Leland, 1974). The standardization populations for the 1969 and 1974 editions were both selected from residential institutions for individuals labeled mentally retarded. The 1969 edition used a standardization population of 2800 residents living in 63 institutions for persons labeled mentally retarded across the United States. The sample was stratified by sex, intelligence, and age of subjects (Miller, 1972). The norming population for the 1974 edition of the ABS consisted of an additional 1200 subjects; bringing the total norming population to 4,000 subjects (Bortner, 1978). A public school version of the ABS was also developed and published in 1975 (Lambert, Windmiller,

Cole & Figueroa, 1975).

There are few instruments which claim to measure adaptive behavior that have normative data obtained from school aged students (Mealor & Richmond, 1980). The A.A.M.D. Adaptive Behavior Scale, Public School Version developed by Lambert, Windmiller, and Cole (1974) is one of the most widely used instruments for assessing adaptive behavior in the public school setting (Mealor & Richmond, 1980) and is normed on school aged students. The development of the instrument occurred because in the words of the authors:

> . . . it has become increasingly apparent that traditional tests of intelligence used alone do not indicate how well a child may function socially nor how well he/she copes with the various demands made upon him/her by the people and the circumstances in his/her environment. Clearly, some measure of a child's ability to engage in social activities and to perform everyday tasks of living has been needed--some measure of his/her adaptive behavior. Before the present standardization project there was no systematic, valid, and reliable way to evaluate the adaptive behavior of noninstitutionalized mentally retarded

children (p. 1).

The ABS-PSV was developed using a pilot study in which the investigators used the ABS (Nihira, Foster, Shellhaas & Leland, 1969) to assess the adaptive behavior of a sample population of 60 third and sixth grade students in regular, EMR and EH classrooms. The goals of the pilot study were to determine: (a) whether the ABS would have enough discriminatory power to obtain differences in the functioning of children assigned to regular and special education programs; (b) whether there would be significant differences between parent and teacher ratings of children's adaptive behavior; and (c) whether there were any differences attributable to ethnic status (Lambert, Windmiller, Cole & Figueroa, 1975, p. 35). The results of the pilot study indicated that the ABS had enough discriminatory power to be an effective instrument for use in the public school setting. Teacher and parent ratings were found to be similar when domain scores were used to determine inter-rater reliability. The ethnic status of the subject had a minimal effect on his/her rating (Lambert, Windmiller, Cole & Figueroa, 1975). The investigators concluded from this pilot study that the ABS could be used in the public school setting.

A standardization project was then undertaken involving 2600 students in the California public schools. The sample was chosen on the basis of six factors: class placement,

sex, density of population in the area of resident, socioeconomic status, ethnic status, and age. From the data collected the investigators were also able to do an item analysis of teachers' comments on the items of the ABS. The teachers' comments were compared to those made by staff members of two unnamed Education, Guidance, and Research staffs in large urban school districts. The result was the A.A.M.D. Adaptive Behavior Scale, Public School Version (Lambert, Windmiller, & Cole, 1974).

The ABS-PSV is divided into two major parts: Adaptive and Maladaptive Behavior. These areas are further divided into 24 behavior domains. Part I of the ABS-PSV consists of the behavior domains labeled Independent Functioning, Physical Development, Economic Activity, Language Development, Numbers and Time, Domestic Activity, Self-Direction and Socialization. The purpose of Part I of the ABS-PSV, according to Nihira, Foster, Shellhaas, and Leland (1974), "is to evaluate an individual's skills and habits in ten behavior domains (coherent groups of related activities) considered important to the development of personal independence in daily living" (p. 1). Part II of the ABS-PSV is "designed to provide measures of maladaptive behavior related to personality and behavior disorders" (p. 1). The behavior domains of Violent and Destructive Behavior, Antisocial Behavior, Rebellious Behavior,

Untrustworthy Behavior, Withdrawal, Stereotyped Behavior and Odd Mannerisms, Inappropriate Interpersonal Manners, Unacceptable or Eccentric Habits, Self-Abusive Behavior, Hyperactive Tendencies, Sexually Abberent Behavior, Psychological Disturbance, and Use of Medications make up the second part of the ABS-PSV.

Administration of the ABS-PSV can take place in three ways: first party assessment, third party assessment, and interview method. First person assessment is recommended by Nihira, Shellhaas, and Leland (1974) if the rater is familiar with the subject being assessed and is able to "judge the relevance of the Scale items" (p. 3). If first person assessment is used within the school setting, the teacher is the person most qualified to use the ABS-PSV. Third party assessment is used when very detailed information is needed and the person who has the most knowledge about the subject is untrained in the use of the ABS-PSV. The interview method of administering the ABS-PSV requires that a person who is very familiar with the content and order of the Scale be the administrator. General open-ended questions are asked by the interview about the subject's behavior as it relates to his/her functioning in the specific domains of the ABS-PSV.

Revision of the ABS-PSV

Recently, the ABS-PSV was revised by Lambert and Windmiller (1981). This revision is entitled the

A.A.M.D. Adaptive Behavior Scale, School Edition (ABS-SE). The ABS-PSV was revised, according to its authors, as a response to criticisms of persons working in the field of mental retardation who asked that scoring procedures be revised and the reference group be expanded to cover a wider range of individuals. Several revisions of the ABS-PSV were undertaken to improve the reliability and validity of the scale. The standardization population was increased from 2600 to 6500 subjects and norms were developed for individuals classified as regular, mildly mentally retarded and moderately mentally retarded. Norms for Part II of the ABS-PSV were redesigned to be consistent with those of Part I. Scoring procedures were simplified and modified so that scaled scores could be obtained from the factor scores. Additional studies were conducted to obtain more accurate information on the reliability and validity of the Scale. The names of two domains were changed and a descriptive guide for parents about the ABS-SE was developed.

The ABS-SE consists of two parts. Part I is organized along developmental lines and is designed to evaluate a person's skills and habits in nine behavior domains including: Independent Functioning, Physical Development, Economic Activity, Language Development,

Numbers and Time, Prevocational Activity, Self-Direction, Responsibility, and Socialization. Part II of the ABS-SE consists of twelve behavior domains labeled: Aggressiveness, Antisocial vs. Social Behavior, Rebelliousness, Trustworthiness, Withdrawal vs. Involvement, Mannerisms, Interpersonal Manners, Acceptability of Vocal Habits, Acceptability of Habits, Activity Level, Symptomatic Behavior, and Use of Medications. The purpose of Part II is to provide a measure of adaptive behavior related to personality and behavior disorders.

There were five sources from which the standardization sample was drawn: (a) the original California sample of the ABS-PSV, (b) a Florida field-test, (c) a California Florida preschool sample, (d) a California reevaluation sample, and (e) a California secondary sample. A total of 6500 persons ranging in ages from three to sixteen was used to generate scaled score and percentile norms. There were 2135 subjects labeled regular, EMR, and TMR, ranging in ages from 6 to 12 years of age, selected from the California ABS-PSV sample. A total of 3220 students, ages 3 to 16, identified as regular, EMR, and TMR, were used in the Florida sample. The California/Florida preschool sample consisted of 363 subjects from 3 to 6

years of age. The fourth source of data was a California reevaluation study which consisted of 656 subjects identified as EMR and TMR. An additional 149 subjects, aged 13 to 16, identified as regular, and EMR were also included in the standardization project. This standardization project was undertaken to expand the use of the ABS-PSV in the public school setting (Lambert & Windmiller, 1981).

The evolution of the A.A.M.D. Adaptive Behavior Scales has shown much effort on the part of its authors to develop a valid and reliable instrument which measures adaptive behavior. Over a decade of research has been conducted by various investigators to achieve this goal. The A.A.M.D. Scales have become some of the most widely used instruments for assessing adaptive behavior in the United States (Mealor & Richmond, 1980; Spreat 1982). One question that remains to be answered about the various editions of the ABS is how reliable the ratings of a subject's adaptive behavior is between raters. The inter-rater reliability of any instrument must be investigated under the mandate of the American Psychological Association (A.P.A., 1974).

Inter-Rater Reliability and Testing

Minimum standards for the development and use of psychological tests have been promulgated by the

American Psychological Association (A.P.A., 1974). These standards include the conducting and the reporting of research related to consistency between scorers or raters. Section F1.1 of the Standards for Educational and Psychological Tests and Manuals (A.P.A., 1974) states:

> The test manual should furnish, insofar as feasible, a quantitative analysis of the total inconsistency of measurement into its major identifiable components; namely inconsistency in responses of the subject; inconsistency or heterogeneity within the sample of test content (such as the stimulus items, questions, and situations); inconsistency among scorers, raters, or units of apparatus; and mechanical errors of scoring (p. 50).

The manual for a test should also state the assessment technique which assures the test user that "tester-effect" is minimized or that has the highest reliability of assessment across raters or scorers (A.P.A., 1974). Hopkins and Hermann (1977) agree with this mandate and extend it to include the calculation and the publication of indices of random-chance inter-rater agreement.

Methods for Determining Inter-Rater Reliability

Inter-rater reliability is defined as the extent to which two observers agree on the rating or scoring of a behavior exhibited by a subject. A high rate of agreement between the two raters would indicate that the observations by the raters reflect the subject's performance accurately (Kazdin, 1977). There are several methods or techniques of determining inter-rater reliability, but the purpose of any technique used is to reflect accurately, with minimum ambiguity, the degree of reliability of the data assessed (Hartmann, 1977).

One of the most commonly used methods of determining inter-rater reliability is the percent agreement technique. The calculation of inter-rater reliability using the percent agreement formula is as follows:

$$\frac{\text{number of agreements}}{\text{number of agreements + number of disagreements}} \times 100$$

Yelton, Wildman, and Erickson (1977) state that the weaknesses of this particular method of calculating inter-rater reliability include: (a) percent agreement is directly affected by the frequency of behaviors; (b) the decision whether or not to include agreement on nonoccurrence of behavior can drastically affect

obtained agreement percentages; and (c) similar agreement percentages do not reflect the same quality of observer agreement because the number of agreements that may have occurred by chance are not considered. Kratochwill and Wetzel (1977) believe that the major problems with using this conventional method of determining inter-rater reliability are: (a) the implication that it cannot always provide a safeguard against misrepresentation of experimental findings and (b) that it does not take into account chance agreements between observers. These and other criticisms have led to the development of more sophisticated statistical procedures for determining inter-rater reliability. These new techniques are often referred to as correlational statistics.

The Kappa coefficient (Cohen, 1960) is an example of a correlational statistic. It was developed to measure the inter-rater reliability of categorical data, accounting for chance agreements between raters. It is calculated using the formula below:

$$\frac{(Po - Pe)}{(1 - Pe)}$$

The quantity (Po - Pe) represents the proportion of observed agreements minus the proportion of chance or

expected agreements. This quantity is divided by the quantity (1 - Pe) which is also corrected for chance agreements between raters.

The maximum for Kappa is + 1.0 when no disagreements are present and both observers exhibit variation in the scoring categories. Kappa will equal zero when the proportion of chance agreements equals the number of observed agreements. Kappa will have a negative value if the proportion of observed agreements is less than the proportion of expected or chance agreements (Hartmann, 1977).

The use of correlational statistics in determining inter-rater reliability has several advantages over conventional methods. Hartmann (1977) lists two such advantages: (a) they are easily calculated and readily interpreted and (b) they can be readily tested for significance. Kratochwill and Wetzel (1977) state that the reporting of correlational statistics greatly improves the statistical judgement and that such statistics decrease subjectivity and the degree of inference in judging agreement scores over percent statistics.

There are some disadvantages of using correlational statistics, Hartmann (1977) states that random factors are more likely to produce low reliability estimates

than systematic factors. Kratochwill and Wetzel (1977) believe that the development of correlational statistics will lead experimenters to use them because of their novelty and exclude simpler techniques which require a less sophisticated knowledge of statistical analysis.

Inter-Rater Reliability Studies Related to ABS School Scales

Lambert, Windmiller, Cole, and Figueroa (1975) state "we did not conduct reliability studies with the public school version . . . ." (p. 51). A review of the literature reveals that only four studies have been conducted on the inter-rater reliability of the ABS-PSV (Brulle & Hoernicke, 1984; Givens, 1980; Mayfield, Forman, & Nagle, 1984; Mealor & Richmond, 1980). The studies conducted by Givens (1980), Mayfield, Forman, and Nagle (1984), and Mealor and Richmond (1980) only investigated the inter-rater reliability of Part I of the ABS-PSV. There have been no published studies concerning the inter-rater reliability of the ABS-SE.

Givens (1980) used two groups of 6 subjects selected at random from graduate students in Education and Special Education as raters for her study. Group one was not given training in the scoring of the ABS-PSV. Group two was given inservice training in the proper scoring techniques for the ABS-PSV. Using a coefficient of agreement, Givens found a mean

inter-rater reliability coefficient of .55 between the two groups. Domain inter-rater reliability coefficients ranged from -.73 to .80. She states that "for an objectively scored instrument, the reliability of the scorers proved to be disappointingly low" (p. 36). Although this study shows that there are significant differences in the ratings of trained and untrained raters, the difference among trained raters is also significant. The coefficient of agreement among the trained observers was only .77; a level that is unacceptable by current testing standards for this measure of inter-rater reliability.

In a study conducted by Mealor and Richmond (1980) using teachers and parents as raters on Part I of the ABS-PSV, significant differences in ratings were also found. The authors asked parents and teachers of students identified as moderately or severely retarded to rate the behavior of their student/child. A Hotelling T2 was used to analyze the data collected. The results revealed that parents' and teachers' ratings differed significantly on five of the ten behavior domains.

Mayfield, Forman, and Nagle (1984) used 31 students enrolled in resource programs for individuals identified as mildly mentally retarded as subjects. Four group

raters, using Part I of the ABS-PSV, were asked to rate the subjects. These groups consisted of parents (mothers), special educators, regular educators, and independent observers. Significant differences in ratings were indicated by a one-way multivariate analysis of variance. Six of the ten behavior domain ratings on Part I of the ABS-PSV resulted in significant discrepancies. A Pearson product-moment correlation was also conducted to determine the relationship of the ratings of the four groups of raters. The mean correlation between parents and regular educators was .67, between parents and special educators was .50, between regular educators and special educators was .52, between regular educators and observers was .50, and between special educators and observers was .45. Disagreement between raters was deemed to be significant by the experimenters with the observation that parents' ratings compared more favorably with the regular educators' ratings than with the other two rating groups.

The most recent study available on the inter-rater reliability of the ABS-PSV was conducted by Brulle and Hoernicke (1984). In this study the authors used teachers who had "extensive day-to-day contact" with the students they were asked to rate. After the data were

collected, it was analyzed using three measures of inter-rater reliability: simple agreement, Cohen's Kappa, (Cohen, 1960), and a weighted average technique (Harris & Lahey, 1978). The simple agreement technique yielded acceptable levels of inter-rater agreement on all of the behavior domains of Parts I and II of the ABS-PSV. When more stringent measures of inter-rater reliability were applied, the inter-rater agreement was dramatically reduced. The Cohen's Kappa produced reliability ratings of -.05 to .67. The weighted average formula resulted in low levels of reliability as well. Ratings of reliability ranged from .02 to .68. The authors state that the reliability coefficients for the overall domains are acceptable for making placement decisions, but that the item-by-item analysis reveals unacceptable coefficients to yield useful information for placement or programming decision makers. This study is limited in its generalizability, however, because the subjects used for the study were all above the suggested age range of the instrument and the sample for the study consisted of only five subjects.

Clearly, the literature has shown the need for an extensive investigation of the inter-rater reliability of the ABS-SE. This investigation should include the use of both Parts I and II of the ABS-SE and the use of

trained raters to afford the instrument a fair evaluation. Analysis of the data collected should be conducted through the use of correlational statistics because ratings on the ABS-SE often determine a student's educational placement and program. The present study systematically addresses each of the needs listed above. The purpose of this study is twofold: (a) to determine an inter-rater reliability coefficient for the ABS-SE using secondary special education personnel as raters and (b) to determine if there are any significant differences in the perceptions of secondary special education personnel related to the adaptive behavior of students identified as mildly mentally retarded as measured by the ABS-SE.

Method

Subjects

Subjects for this study were randomly selected from a list of students enrolled in a secondary (9-12) program for individuals identified as mildly mentally retarded. The subjects attended a public high school with a student population of approximately 2150; located in a community of approximately 40,000 residents. The subjects ranged in age from 14 to 18 years old and IQ scores ranged from 75 to 55. The ethnic status of the subjects was divided almost equally into two groups:

Black and Caucasian. The original sample consisted of 10 males and 10 females for a total of 20 subjects. One male subject and one female subject were dropped from the sample because raters who had initially indicated that they could rate the subjects using the ABS-SE felt that their ratings would be invalid on the basis of a lack of information. The subjects could not be replaced due to time limitations; therefore, the total sample consisted of 18 subjects (9 males and 9 females).

Raters

Raters were all secondary (9-12) special education personnel with certification and teaching experience in the area of mild mental retardation. All raters were part of the support services staff of a midwestern high school which employs a total teaching and administrative staff of 125 members. There were four female raters and 1 male rater; all of which were Caucasian. Raters of a particular subject were randomly selected as described in the procedures section of this study. All raters had either daily contact or sufficient information to rate the subjects to which they were assigned.

Procedures

There were three major components in the implementation of this investigation: (a) subject selection, (b) rater selection, and (c) rater inservice.

Subjects were randomly selected from a list of students enrolled in a secondary (9-12) program for students identified as mildly mentally retarded. After the subjects were selected, a list was circulated to raters containing the names of the subjects. Potential raters were asked to indicate which students they had daily contact with or had enough information that they could complete the ABS-SE using the student as a subject. Two raters were then randomly selected from the list of potential raters who had indicated that they could rate the behavior of a subject. All of the raters had previous training in the administration of the ABS-SE. An individual training session reviewing the proper scoring and administration of the ABS-SE was provided by the investigator to all raters. Raters had access to the investigator on matters of a technical nature, but no other help was given by the investigator. Raters were asked to complete the ABS-SE assessment booklet independently and not to discuss their ratings. The assessment booklets were then returned to the investigator for analysis.

Results

Organization Pattern of ABS-SE

The ABS-SE consists of two major components Part I, which measures adaptive behavior and Part II, which

measures maladaptive behavior. Each part consists of domains and subdomains (Part I only). Domains and subdomains are made up of items which assess behaviors related to the particular domain or subdomain.

Analysis of Data

The data collected were analyzed using two methods of calculating inter-rater reliability: (a) overall reliability based on total domain or subdomain scores [the procedure most commonly used, according to Hartmann (1977)] and (b) Cohen's kappa (Cohen, 1960). Cohen's kappa allows for a stringent item-by-item analysis by correcting for chance agreements. Harmann (1977) states that a reliability coefficient of .80 is required to be considered acceptable by many researchers when the percentage agreement formula is used to calculate inter-rater reliability. A kappa of .60 or higher is recommended by Gelfand and Hartman (1975) to be considered acceptable. The overall reliabilities of the subdomains and domains were calculated by dividing the agreements, on both occurrence and nonoccurrence, by the total number of agreements and disagreements. The comparison of the ratings using kappa provided for a more complex analysis of the data. Part I of the ABS-SE consists of two types of items: those which ask the rater to circle only one response and those items which

ask the rater to circle all statements which apply to the individual being rated. See Appendix I for sample items.

For those items which asked the rater to circle only one response for a certain point total, agreements on occurrence were scored for all statements less than or equal to the total circled by each rater. Agreements on nonoccurrence were scored for statements with a point value greater than the total. If an item had a total point value of "5" and rater 1 circled a "4" and rater 2 a "3", three agreements on occurrence were scored (statements 1-3), one agreement on nonoccurrence was scored (statement 5), and one disagreement (statement 4, rater 1 - occurrence, rater 2 - nonoccurrence) was scored. On those items which asked the rater to circle all statements which applied, an agreement on occurrence was scored if each rater circled the statement and an agreement on nonoccurrence was scored if neither rater circled the statement (Brulle & Hoernicke, 1984). On Part II of the ABS-SE, the rater is given a statement and asked to rate the occurrence of the particular behavior as "occasionally" or "frequently". If the particular behavior does not occur neither descriptor is circled. See Appendix I for sample item.

Each statement was considered two separate

statements for the purpose of this investigation. If both raters indicated that a particular behavior occurred "occasionally", an agreement on occurrence was scored. If neither rater indicated that a particular behavior occurred "occasionally", an agreement on nonoccurrence was scored. If either rater indicated a particular behavior occurred "occasionally" and the second rater did not indicate such, a disagreement was scored. This process was then repeated for the statement using the descriptor "frequently". A 2 x 2 matrix, having the following cells: agreement on occurrence; agreement on nonoccurrence; disagreement [rater 1 - occurrence and rater 2 - nonoccurrence]; and disagreement [rater 1 - nonoccurrence and rater 2 - occurrence], was then prepared to summarize the data for each subdomain, domain, and part of the ABS-SE.

The results of both techniques of estimating inter-rater reliability are listed in Table 1. Scores in the first column indicate inter-rater reliability estimates of overall factors only. The second column shows the inter-rater reliability estimates when a more stringent correlational statistic is used for item-by-item agreement analysis. The estimates of inter-rater reliability using kappa were somewhat lower than those using calculated overall factor scores. This

is because of the statistical nature of kappa and was an expected result.

Table 1

Reliability Estimates

Component	Reliability	Kappa
Independent Functioning	.92	.82
Eating	.91	.82
Toilet Use	.98	-.01
Cleanliness	.96	.90
Appearence	.78	.45
Care of Clothing	.74	.43
Travel	.99	.93
Other Independent Functioning	.88	.68
Physical Development	.94	.76
Sensory Development	.92	.58
Motor Development	.94	.80
Economic Activity	.79	.55
Money Handling & Budgeting	.73	.38
Shopping Skills	.85	.69
Language Development	.85	.65
Expression	.87	.72
Comprehension	.85	.52
Social Language Development	.80	.53

Table 1 (continued)

Reliability Estimates

Component	Reliability	Kappa
Numbers and Time	.92	.77
Prevocational Activity	.75	.49
Self-Direction	.76	.52
Initiative	.78	.53
Perserverence	.77	.53
Leisure Time	.68	.36
Socialization	.82	.63
Responsibility	.79	.49
Overall Part I	.87	.70
Aggressiveness	.92	.70
Anti-Social vs. Social Behavior	.87	.47
Rebelliousness	.85	.46
Trustworthiness	.87	.53

Table 1 (continued)

Reliability Estimates

Component	Reliability	Kappa
Withdrawal vs. Involvement	.92	.67
Mannerisms	.97	.71
Appropriateness of Interpersonal Manners	.94	.71
Acceptability of Vocal Habits	.86	.30
Acceptability of Habits	.98	.92
Activity Level	.86	.47
Symptomatic Behavior	.88	.47
Use of Medications	.96	.84
Overall Part II	.90	.59

Discussion

This study has some limitations in its generalizability to the overall population for which the ABS-SE is used. The sample population is small and demographically restricted in terms of the age range and area of residence of the subjects. Several interesting findings, however, were obtained and warrant further discussion. The inter-rater reliability estimates using only the domain or subdomain scores are quite high with the majority of the reliability estimates exceeding .80. This finding supports that of Brulle and Hoernicke (1984) and also their contention that the ABS-PSV can be useful in making placement decisions, because the items on the ABS-PSV and the ABS-SE are the same.

The relatively high scores obtained when kappa was used to analyze data indicates that the ABS-SE may be useful in making programming decisions for some aspects of a subject's adaptive behavior. The majority of the subdomains and domains on Part I of the ABS-SE had acceptable levels of inter-rater reliability. The subdomains and domains of Travel, Sensory Development, Economic Activity, Comprehension, Social Language Development, and Self Direction are questionable as to their usefulness in aiding diagnosticians and educators in making programming decisions. The reliability

rating of .59 indicates that overall, Part II of the Scale is slightly below the accepted levels of inter-rater reliability necessary for making such decisions. The domain of Trustworthiness is also questionable as to its usefulness in providing information relevant to programming. The domains of Antisocial vs. Social Behavior, Rebelliousness, Acceptability of Vocal Habits, and Symptomatic Behavior provide programming decisions makers with little or no reliable information.

Low coefficients of reliability between raters on the components of Travel, Economic Activity, Leisure Time, Money Handling and Budgeting, and Care of Clothing may be attributed to the lack of direct observation of the particular skills by the raters. The use of public transportation, banking facilities, and budgets are skills that a classroom teacher may never observe directly unless the curriculum calls for the assessment and teaching of such skills. The inter-rater reliability of such items might be higher if direct observation of the skills being rated occurred more often by the raters.

Each of the components in Part I of the ABS-SE which have questionable or unacceptable levels of inter-rater reliability consist of only one or two

items. This lack of information being requested about a certain set of skills coupled with the lack of direct observation of the skills on which information is being requested, could be a reason why such low reliability coefficients were obtained. If the components were expanded in such a way that there were more specific items which requested information that teachers can observe directly, the reliability coefficients may improve.

The domains on Part II of the ABS-SE which have low inter-rater reliability share the problems of those components on Part I of the scale which have unacceptable reliability coefficients and have some problems that are unique to this particular part of the scale. The domains of Anti-Social vs. Social Behavior, Rebelliousness, Acceptability of Vocal Habits, Activity Level, and Symptomatic Behavior all ask the rater to rate behaviors which every child is going to exhibit at some time or another during his/her school career. The behaviors may be more intense in one teacher's classroom than another's and teachers' personalities and toleration for certain behaviors are very different. A student's age, the time of day, and many other factors may cause a student to display the behaviors being rated with a high degree of frequency at a particular time,

but he/she may not display the behaviors at another given time. Another problem with the domains on Part II of the ABS-SE is that they often ask the rater to rate behaviors which they have not been trained to observe. The domain of Symptomatic Behavior is an example of this phenomenon. Items include questions about the student's feelings of persecution, hypochondrical tendencies, and signs of emotional instabilities. Unless the rater is trained in the observation of such behaviors and abnormal psychology, the ratings given cannot be valid nor reliable.

Each of the components that have low kappa ratings have some common characteristics which, if changed, might make them more reliable and useful in program decision making. The wording of the items, or statements within items, are vague and leave the rater guessing as to the information being asked for in the questions (Elliot, 1985). Few statements or items are stated in observable or measureable terms. Words such as usually, adequately and obviously, used in items and statements within items, ask the rater to make judgements which result in a myriad of responses. Items on Part II of the ABS-SE also possess this characteristic of vagueness when raters are asked to rate the occurrence of a particular behavior as

"occasionally" or "frequently". The frequency of a behavior should be rated and expressed in terms of occurrence per minute, per hour, per day, etc. . . A measure of intensity would also make the information obtained more valuable to the diagnostician and the educator.

Raters in this investigation were asked about particular problems, if any, they encountered while rating the adaptive behavior of subjects. They voiced several concerns. Raters stated that items on Part I of the ABS-SE switch frequently in type of response that is required of the rater (circle only one or circle all that apply). A rearrangement of items so that those which require a similar type of response are grouped together within a particular subdomain or domain would make the rating procedure less confusing. Another problem with the items on the ABS-SE, according to raters in this study is that statements within items switch from being phrased in the positive to being phrased in the negative; thus causing the rater to become confused about the type of response that is being asked for by the item. Simply stating all questions of a particular item in the same manner (positively or negatively) would rectify this problem and lead to greater reliability between raters because

of increased clarity of the items. The above suggestions represent what the investigator considers to be changes that will increase the inter-rater reliability of the ABS-SE and thus increase the instrument's usefulness as a programming tool.

This investigation, even with its limitations, has produced results tht are relevant and important to the measurement of adaptive behavior. First, the study has shown that the ABS-SE can be a reliable instrument for school personnel making placement decisions. The investigation has shown that the ABS-SE may have some usefulness as a programming tool. The majority of the items on the ABS-SE withstood the scrutiny of a very stringent measure of inter-rater reliability. This allows diagnosticians and educators assurance of accurate information which is vital to the education process. The contention that teachers are the most qualified to rate the behavior of students on the ABS-SE by its authors was also supported by the reliability coefficients obtained through the analysis of the data. Further investigation of the inter-rater reliability of the ABS-SE is warranted on the basis of the findings of this study to provide more accurate information about the usefulness of the ABS-SE as a measure of adaptive behavior.

Several unanswered questions have also been raised by this investigation. Would more reliable data be produced if both the teacher and parent were used as raters? Would a different technique of completing the ABS-SE, such as third party assessment or interview, produce more reliable results than the use of trained raters using the first party assessment technique? Would ratings between teachers on the ABS-SE be more reliable if the student has been identified as moderately, severely, or profoundly mentally retarded? These are questions that need to be addressed by further research. We (researchers, educators, diagnosticians, and parents) must continue to seek out and refine the measurement of this concept (adaptive behavior) which has been present from the earliest definitions of mental retardation, but has eluded reliable and valid measurement to the present day.

References

American Psychological Association. (1974). Standards for educational and psychological tests and manuals. Washington, D.C.: American Psychological Association.

Bortner, M. (1978). AAMD Adaptive Behavior Scales. In O.K. Buros (Ed.). Eighth mental measurement yearbook (pp. 492-493). Highland Park, NJ: The Gryphon Press.

Brulle, A. & Hoernicke, P. (1984). A brief report on the reliability of the public school version of the A.A.M.D. Adaptive Behavior Scale. The Mental Retardation and Learning Disability Bulletin, 12(2), 115-118.

Campbell, I. (1985). Vineland Adaptive Behavior Scales. In J. V. Mitchell (Ed.). Ninth mental measurement yearbook (pp. 1660-1662). Highland Park, NJ: The Gryphon Press.

Clausen J. (1972). Quo vadis, AAMD? Journal of Special Education, 6(1), 51-60.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37-46.

Diana vs. California State Board of Education. Civ. Act. No. 70-37 RFP. (N.D. Cal. Jan. 7, 1970 & June 18, 1973).

Doll, E. (1924). Capabilities of low grade feeble-minded. Training School Buletin, 21, 65-77.

Doll, E. (1929). Community control of the feebleminded. Journal of Psycho-Asthenics, 34, 161-175.

Doll, E. (1936). The Vineland Social Maturity Scale. Publication of the Training School at Vineland, Department of Research, Series 1936, No. 3.

Dunn, L. (1968). Special education for the mildly mentally retarded - Is much of it justifiable? Exceptional Children, 35, 5-22.

Elliot, S. (1985). A.A.M.D. Adaptive Behavior Scale (School Edition). In J.V. Mitchell (ed.). Ninth mental measurement yearbook (pp. 1-4). Highland park, NJ: The Gryphon Press.

Garrison, I. (1960). Developing potential of exceptional children. Exceptional Children, 26, 510.

Gelfund, D. & Hartmann. (1975). Child behavior analysis and therapy. New York: Pergamon Press.

Givens, T. (1980). Scorer reliability on the A.A.M.D. adaptive behavior scale, public school version, part one. Psychology in the School, 17, 335-338.

Goodenough, F. (1949). Mental testing. New York: Rinehart.

Grossman, H. (Ed.) (1973). Manual on terminology and classification in mental retardation. Washington, D.C.: American Association on Mental Deficiency.

Harris, F. & Lahey B. (1978). A method of combining occurrence and nonoccurrence interobserver agreement scores. Journal of Applied Behavior Analysis, 11, 523-527.

Hartmann, D. (1977). Considerations in the choice of interobserver reliability estimates. Journal of Applied Behavior Analysis, 10 (1), 103-116.

Herber, R. (ed.) (1961) Manual on terminology and classification in mental retardation. Washington, D.C.: American Association on Mental Deficiency.

Hopkins, B. & Herman, J. (1977). Evaluating interobserver reliability of interval data. Journal of Applied Behavior Analysis, 10(1), 121-126.

Kazdin, A. (1977). Artifact, bias, and complexity of assessment: The a b c s of reliability. Journal of Applied Behavior Analysis, 10(1), 141-150.

Kratochwill, T. & Wetzel, R. (1977). Observer agreement, credibility, and judgement: Some considerations in presenting observer agreement data. Journal of Applied Behavior Analysis, 10(1), 133-139.

Lambert, N. & Windmiller M. (1981). A.A.M.D. Adaptive Behavior Scale, School Edition. Washington, D.C.: American Association on Mental Deficiency.

Lambert, N., Windmiller, M. & Cole, E. (1974). A.A.M.D. Adaptive Behavior Scale. Washington D.C.: American Association on Mental Deficiency.

Lambert, N., Windmiller, M., Cole, E. & Figueroa, R. (1975). Manual of the A.A.M.D. Adaptive Behavior Scale, Public School Version. Washington D.C.: American Association on Mental Deficiency.

Larry P. vs. Wilson Riles. 343 F. Supp. 1306, N.D. Cal. 1972, 502 F. 2d. (9th Cir. 1974).

Lord, F. (1956). A realistic look at special classes. Exceptional Children, 22, 321-325, 342.

Mayfield, K., Forman, S., & Nagle, R. (1984). Reliability of the A.A.M.D. Adaptive Behavior Scale-Public School Version. The Journal of School Psychology, 22, 53-61.

Mealor, D. & Richmond, B. (1980). Adaptive behavior: Parents and teachers disagree. Exceptional Children, 46(5), 386-389.

Miller, L. (1972). AAMD Adaptive Behavior Scales. In O.K. Buros (Ed.) Seventh mental measurement yearbook. (pp. 68-70). Highland Park, NJ: The Gryphon Press.

Nihira, K. Foster, R., Shellhaas, M. & Leland, H. (1967). A.A.M.D. Adaptive Behavior Checklist. Washington, D.C.: American Association on Mental Deficiency.

Nihira, K., Foster R., Shellhaas, M. & Leland, H. (1969). A.A.M.D. Adaptive Behavior Scale. Washington, D.C.: American Association on Mental Deficiency.

Nihira, K., Foster, R., Shellhaas, M. & Leland, H. (1974). A.A.M.D. Adaptive Behavior Scale. Washington, D.C.: American Association on Mental Deficiency.

Nihira, K. & Nihira, L. (1975). Jeopardy in community placement. American Journal on Mental Deficiency, 79(5), 538-544.

Payne, J., Polloway, E., Smith, J. & Payne, R. (1981). Strategies for teaching the mentally retarded. Columbus, OH: Charles E. Merrill Publishing Company.

Porteus, S. (1921). A social rating scale for defectives. Journal of Psycho-Asthenics, 26, 117-126.

President's Panel on Mental Retardation (1962). A proposed program for national action to combat mental retardation. Washington, D.C.: Superintendent of Documents.

Salvia, J. & Ysseldyke, J. (1978). Assessment in remedial and special education. Boston: Houghton-Mifflin.

Scheerenberger, R. (1983). A history of mental retardation. Baltimore, MD: Brookes Publishing Company.

Spreat, S. (1982). The AAMD Adaptvie Behavior Scale: A psychometric review. The Journal of School Psychology, 20(1), 45-56.

Yelton, A., Wildman, B. & Erickson, M. (1977). A probability-based formula for calculating interobserver agreement. Journal of Applied Behavior Analysis, 10(1), 127-131.

Yepsen, L. (1941). Defining mental deficiency. American Journal of Mental Deficiency, 46, 200-205.

Appendix 1

CARE OF CLOTHING SUBDOMAIN

ITEM 11 Care of Clothing

(Circle all that apply)

Wipes and polishes shoes when needed	1
Puts clothes away neatly	1
Hangs up clothes without prompting	1
Does none of the above	0
SUBDOMAIN TOTAL (Enter Item 11)	

DRESSING AND UNDRESSING SUBDOMAIN

ITEM 12

Dressing

(Circle only one)

Dresses self completely (raincoat, overshoes, etc.)	5
Dresses self completely with verbal prompting	4
Dresses self by pulling or putting on all clothes with verbal prompting and by fastening them (zipping, buttoning, snapping) with help	3
Dresses self with help in pulling or putting on most clothes and in fastening them	2
Cooperates when being dressed by extending arms or legs	1
Must be dressed completely	0

Appendix 1 (continued)

ITEM 13	Shoes
	(Circle all that apply)
Puts on shoes correctly without assistance	1
Ties shoelaces without assistance	1
Unties shoelaces without assistance	1
Removes shoes without assistance	1
Does none of the above	0
SUBDOMAIN TOTAL (Add Items 12 & 13)	

ITEM 87	Frequently	
	Occassionally	
HAS HYPERACTIVE TENDENCIES		
Talks excessively	1	2
Will not sit still for any length of time	1	2
Constantly runs or jumps around the room	1	2
Moves or fidgets constantly	1	2
Does none of the above	0	0
Other (specify)	1	2